From:	PETERSON Jenn L
To:	Joe Goulet/R10/USEPA/US@EPA; Burt Shephard/R10/USEPA/US@EPA; Robert.Neely@noaa.gov; Jay.Field@noaa.gov; Eric Blischke/R10/USEPA/US@EPA; rgensemer@parametrix.com; JMalek@parametrix.com; Bob Dexter
Cc:	Chip Humphrey/R10/USEPA/US@EPA
Subject:	Reliability
Date:	06/03/2008 12:22 PM
Attachments:	SQG reliability memo.doc TeresaMDEQ reliabilityAnalysis.doc

---

I am including some memos by DEQ and RSET (Teresa) on these issues. The memos are based on a review of the proposed RSET framework using the floating percentile methodology. RSET terms are used (SL1, SL2), but the reliability concepts are the same. SQG determinations based on a regional FPM model was proposed in this case was to be the only line of evidence in evaluating the potential toxicity of sediment. Selection of reliability criteria attributes a level of confidence of the model predictions. DEQ has advocated using the no-hit sensitivity (see memo) as the criteria that would need evaluation for the use a predictive model. Ideally, the predicted no-hit sensitivity would be above 90%, and we would accept that we could be wrong 10% of the time (the model predicted it to be non-toxic, when in actuality it was toxic). We have talked with LWG about this issue, and it was my understanding that this measure would be presented and used in model decision making. For the FPM, this is the only measure of "confidence bounds" for the model. Also, reliability in some areas may increase for important measures as you have more data to define the range of effects. That is why RSET is now compiling many more regional datasets than they had originally in order to develop the new set of SQGs, and why we may want to look beyond just site-specific data for the model. Reliance on a site-specific model only may result in low reliability / confidence in the use of the model output (esp. for some important measures); , and we therefore need to balance relevancy and confidence in model predictions appropriately. We should consider combining /using the larger regional dataset if we find it increases the reliability of the model.

Also, it should be noted that reliability measures so far have been focused on evaluating the model with the same dataset that was used to develop the model. The best way to evaluate / validate the model is to evaluate the reliability of new data not used in model development. Again, I will advocate that the new Round 3 data for this purpose to evaluate the reliability of the current models before combining Round 2 and 3 to make another "new model".

-Jennifer

-----Original Message-----
From: Goulet.Joe@epamail.epa.gov [mailto:Goulet.Joe@epamail.epa.gov]
Sent: Monday, June 02, 2008 3:58 PM
To: Shephard.Burt@epamail.epa.gov; Robert.Neely@noaa.gov; Jay.Field@noaa.gov; Blischke.Eric@epamail.epa.gov; rgensemer@parametrix.com; PETERSON Jenn L; JMalek@parametrix.com; Bob Dexter
Cc: Humphrey.Chip@epamail.epa.gov
Subject: Call about Hyalella bioassay hit levels for Portland Harbor

All,

Burt has asked me to organize a call to wrap up our position on the hit levels for Portland Harbor. 9am Tuesday, st for most Lets use the TCT call in number, Non- ; pass code Non-

The issue is whether to continue to use the hit levels that we have used so far and are in the problem formulation or modify them based on RESET/LWGs proposal.

Thanks.

Joe